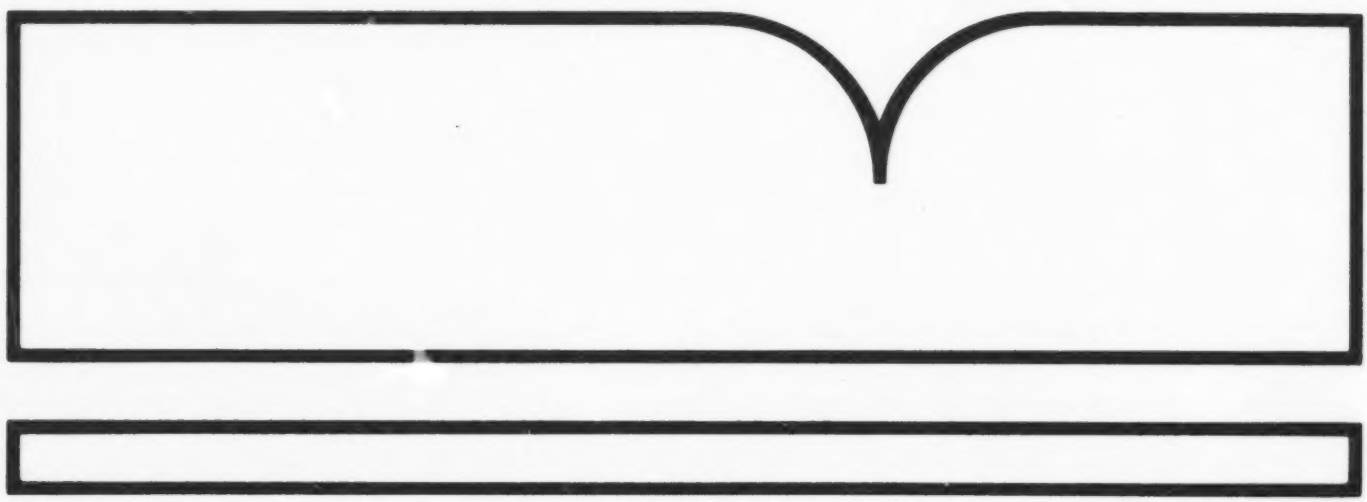
PB90-216094

Principal Component Analysis of SO4(2-) Precipitation Concentrations over the Eastern United States

(U.S.) Environmental Protection Agency, Research Triangle Park, NC

1989



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA/600/J-89/344		PB 90 216094/AS

4. TITLE AND SUBTITLE	5. REPORT DATE
A PRINCIPAL COMPONENT ANALYSIS OF $SO_4^=$ PRECIPITATION CONCENTRATIONS OVER THE EASTERN UNITED STATES	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Brian K. Eder	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Same as Block 12	N104/C/05 5104-05 (FY-89)
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Atmospheric Research and Exposure Assessment Laboratory Office of Research and Development U. S. Environmental Protection Agency Research Triangle Park, NC 27711	14. SPONSORING AGENCY CODE EPA/600/09

15. SUPPLEMENTARY NOTES

To be submitted to Atmospheric Environment

16. ABSTRACT

The spatial and temporal variability of $SO_4^=$ concentrations in precipitation over the eastern United States during the period 1981-1986 was examined through the use of principal component analysis. Application of Kaiser's Varimax orthogonal rotation led to the delineation of seven contiguous subregions, each displaying statistically unique $SO_4^=$ concentration characteristics. These seven statistically significant modes of variability, which together accounted for 74.2% of the total variance, corresponded well with major SO_x emission patterns. Examination of the time series associated with subregion revealed a general seasonality in which periods of high concentrations are more likely during the summer, while periods of low concentrations are more likely during the winter. This seasonal cycle, however, was more prevalent in those subregions which contained few major emissions, and was less prevalent and often obscured by perturbations in those subregions which contained major emissions.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	13
	20. SECURITY CLASS (This page)	22. PRICE
	UNCLASSIFIED	

PB90-216094

Atmospheric Environment Vol. 23, No. 12, pp. 2739–2750, 1989
Printed in Great Britain

Pergamon Press plc

EPA/600/J-89/344

A PRINCIPAL COMPONENT ANALYSIS OF SO_4^{2-} PRECIPITATION CONCENTRATIONS OVER THE EASTERN UNITED STATES

BRIAN K. EDER*

Atmospheric Research and Exposure Assessment Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, U.S.A.

(*First received* 1 *August* 1988 *and in final form* 8 *May* 1989)

Abstract—The spatial and temporal variability of SO_4^{2-} concentrations in precipitation over the eastern U.S. during the period 1981–1986 was examined through the use of principal component analysis. Application of Kaiser's Varimax orthogonal rotation led to the delineation of seven contiguous subregions, each displaying statistically unique SO_4^{2-} concentration characteristics. These seven statistically significant modes of variability, which together accounted for 74.2% of the total variance, corresponded well with major SO_x emission patterns. Examination of the time series associated with each subregion revealed a general seasonality in which periods of high concentrations are more likely during the summer, while periods of low concentration are more likely during the winter. This seasonal cycle, however, was more prevalent in those subregions which contained few major emissions, and was less prevalent and often obscured by perturbations in those subregions which contained major emissions.

Key word index: Principal component analysis, precipitation chemistry, Acid Deposition System, spatial and temporal analysis, SO_x emissions, time series.

INTRODUCTION

The Acid Deposition System (ADS) was established by the Environmental Protection Agency in order to create a long-term, quality precipitation chemistry data base through the integration of wet deposition data from multiple precipitation chemistry networks. Under this system, the SO_4^{2-} concentration data examined in this study, as well as other precipitation chemistry data, have been collected using various sampling protocols at an increasing number of stations during the past decade. Initially, investigators analyzed the resulting data using only rudimentary statistical techniques. More recent endeavors, however, have attempted to analyze the precipitation chemistry data using higher order, more robust statistics, and include the work of Dana and Easter (1987), who used regression techniques, Watson and Olsen (1984a) and Seilkop and Finkelstein (1987), who used kriging techniques and Doty (1986), who used cluster analysis to examine the data. Although such endeavors signify an amelioration in analytical techniques, these methods do not allow the investigator to simultaneously examine the spatial and the temporal variability of precipitation chemistry as does the multivariate technique, Principal Component Analysis (PCA), employed in this study.

Multivariate techniques, such as PCA or Factor Analysis, are not new to the study of air pollution meteorology, or more specifically, precipitation chemistry (Wolff *et al.*, 1986; Gatz, 1978; Hopke *et al.*, 1976; Smeyers-Verbeke *et al.*, 1984; Henry and Hidy, 1979; Thomas, 1986). The primary purpose of PCA in the past, however, has not been to determine spatial and temporal variability of such precipitation, but rather to determine the constituents contributing to the acidity of the precipitation. Such was the case with Crawley and Sievering (1986), who performed factor analysis upon precipitation chemistry data obtained during the late 1970s. Like investigators before them, who limited their studies to one ADS network, Crawley and Sievering were somewhat limited by a dearth of acceptable stations as well as numerous quality-control problems inherent to the early precipitation chemistry data.

In order to circumvent the problems encountered in these early investigations, and so that the characteristics of the spatial and temporal variability of SO_4^{2-} concentration data over the eastern U.S. can be examined simultaneously, PCA has been applied to SO_4^{2-} concentration data obtained from several networks which comprise the ADS for the period from 1981 to 1986. PCA has been employed successfully in the manner suggested in this study by Ashbaugh *et al.* (1984), who used the analysis to examine the spatial and temporal variability of ambient air concentrations of sulfur in eight western states. As employed in this present study, PCA will allow us to spatially delineate areas of homogeneous SO_4^{2-} concentrations, which will help determine which regions within the study areas have similar acid precipitation characteristics.

*On assignment from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

Examination of the time series associated with each of the homogeneous areas will allow analysis and documentation of specific acidic precipitation periods that have occurred in each of the spatially delineated regions.

The sole intent of this paper is to provide a starting point for a variety of interpretive studies by the extended research community. Upon detailed examination and physical interpretation, such analysis will hopefully lend itself to the scientific advancement of acid precipitation.

PRECIPITATION CHEMISTRY DATA

The data employed in this analysis were obtained from three of the networks which comprise the Acid Deposition System (ADS), and include the National Atmospheric Deposition Program (NADP), the Utility Acid Precipitation Study Program (UAPSP) and the Multistate Atmospheric Power Production Pollution Study (MAP3S). In addition to the data screening procedures employed by the individual networks, the ADS has its own program for checking data quality and completeness (Watson and Olsen, 1984b). In order to achieve the optimum number of station-months, the analysis included all available stations having complete records for the longest possible time. This resulted in the selection of 34 stations, as depicted in Fig. 1, which recorded data over the 72 month period from January 1981 to December 1986. A longer period of record could have been examined, but at the expense of the number of available stations and therefore the spatial resolution of the data. Of the 34 sites selected for the study, seven were from MAP3S, four were from UAPSP and 23 were from NADP.

Although the monitoring sites available from the three networks tend to be somewhat irregularly distributed, the sites used in the analysis were selected to be as reasonably close to an equal-area grid as possible. With the exception of a few areas, most notably in the south, the spatial resolution of the combined networks is of the order of 100–200 km. Such resolution sufficiently indicates the large scale spatial patterns of acidic precipitation, which is the aim of this study.

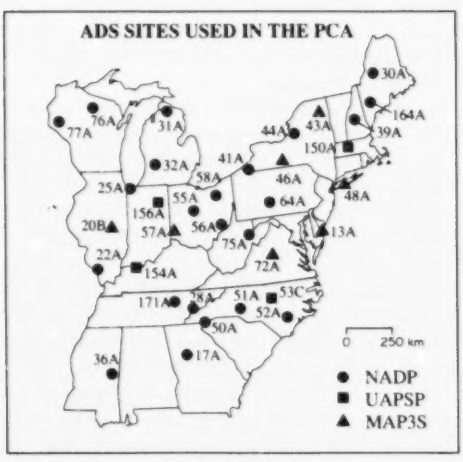


Fig. 1. Acid Deposition System (ADS) sites used in the analysis.

The precipitation chemistry data from each of the three networks were collected using various sampling times, which in the past has somewhat prohibited cross-network analysis. To circumvent this problem, the sampling periods from the three networks, which range from weekly for NADP to daily for UAPSP and event for MAP3S, were converted or 'standardized' into monthly time intervals. These monthly increments represent the precipitation (volume) weighted SO_4^{2-} concentration averages and are computed by the following equation:

$$C = \sum_{i}^{N}\left(C_i * V_i\right) \Big/ \sum_{i}^{N} V_i; \qquad (1)$$

where C_i and V_i correspond to the concentration and precipitation volumes of the N individual samples obtained during the month. N will vary depending upon the network and the meteorological conditions at the specific station. Assuming no sample contamination and total capture, C will then represent the SO_4^{2-} concentrations (μmoles ℓ^{-1}) that would be observed in the sample obtained in a collector left in continuous operation throughout the complete month (MAP3S/RAINE research community, 1982). Computing precipitation volume weighted average concentrations in this manner, equalizes the importance of light rains in which concentrations are generally higher, and heavy rains in which constituents are more dilute (Hendry and Brezonik, 1982).

When incorporating data from several networks in this manner, each of which employed different sampling periods, one must be concerned about the introduction of sampling bias into the resulting data set. Investigation into germane literature revealed inconclusive results as to the possible adverse affects of such an aggregation of data. For example, Sisterson *et al.* (1985), who examined the chemical differences between weekly and event samples collected in northeastern Illinois during the period April 1980–March 1982, found that weekly samples of SO_4^{2-} concentrations averaged generally 8% higher than the daily samples for all seasons except summer and that this difference was more pronounced during the winter. The observed differences were attributed to chemical degradation which occurred in the weekly samples upon waiting for collection and during transport between the field and the laboratory.

Conversely, de Pena *et al.* (1985) found that weekly samples yielded lower SO_4^{2-} concentrations than did event samples recorded at a site near Penn State. In their study, weekly concentrations averaged 8.7% lower than the event samples over a 27 month period.

The lower values obtained from the weekly values were attributed to an increased biological uptake of ions, especially during the summer months when the temperature and therefore biological activity is higher.

Discrepancies such as these may be attributed to the fact that each of these studies were performed at just one site, and therefore the results are specific to that site. It is likely that results from additional studies performed at different locations and climates will vary as well, thereby prohibiting investigators from making adjustments to aggregate data sets. The advantage gained in the spatial coverage of the aggregate data set through the incorporation of different networks, however, far outweighs the disadvantage of the apparently inconsequential bias that may be introduced by incorporating data of varying time periods.

PRINCIPAL COMPONENT ANALYSIS

One of the main objectives of PCA is to identify, through a reduction of data, the recurring and independent modes of variation (signals) within a large, noisy data set, thereby summarizing the essential information of that data set so that meaningful and descriptive conclusions can be made (Hayden and Smith, 1981). The analysis sorts initially correlated data into a hierarchy of statistically independent modes of variation which explain successively less and less of the total variation. PCA can be performed on either covariance, cross-products, or correlation input matrices. Since the identification of spatial modes of variation associated with SO_4^{2-} precipitation was the main objective of this work, a correlation matrix was selected because it is most suitable for resolving spatial oscillations (Overland and Preisendorfer, 1982). Also, selecting a correlation matrix for use in PCA allows isopleths of component loadings (which can be regarded as the correlation coefficient between the component and the individual station) to be drawn.

From the original data matrix, having dimensions of 34 (stations) × 72 (months) and containing 2448 observations, a square, symmetric correlation matrix **R**, having dimensions of 34 × 34 was extracted. By using **R** and the identity matrix (**I**) of the same dimension, the 34 characteristic roots or eigenvalues (λ) were derived, that satisfied the following polynomial equation:

$$\det\left[{}_{34}\mathbf{R}_{34} - \lambda_{34}\mathbf{I}_{34}\right] = 0. \tag{2}$$

For each eigenvalue of Equation (2), which is called the characteristic equation, a non-zero vector **e** can be derived such that:

$${}_{34}\mathbf{R}_{34}\mathbf{e}_1 = \lambda_{34}\mathbf{e}_1, \tag{3}$$

where the vector **e** is called the characteristic vector or the eigenvector of the correlation matrix **R**, associated with its corresponding eigenvalue.

The eigenvectors derived from the correlation matrix represent the mutually orthogonal linear combinations (or modes of variation) of the matrix. Their associated eigenvalues represent the amount of total variance that is explained by each of the eigenvectors. By retaining only the first few eigenvector–eigenvalue pairs, or principal components, a substantial amount of the total variance can be explained while ignoring the higher order principal components which explain minimal amounts of the total variance and can therefore be viewed as noise.

The exact number of principal components that should be retained has been a subject of some controversy and uncertainty. Retaining too few components results in the blending of spatial patterns which are actually more discrete; whereas retaining too many components will delineate areas which represent nothing more than randomly generated variations (Richman, 1981). The method selected for this analysis suggests that all components for which the eigenvalue is $\geqslant 1$ should be retained. The rationale behind this method is that an eigenvalue of 1 represents the amount of variance explained by the original variable, so that a component with an eigenvalue of <1 explains less variance than did any of the original variables (Johnston, 1978). As seen in the next section, use of this eigenvalue $\geqslant 1$ criterion yielded physically interpretable results (i.e. contiguous subregions), therefore inclusion of analysis retaining additional factors did not seem warranted.

The results of using this test, found in Table 1, suggest that the first seven principal components should be retained, and that the higher order principal components, PC 8, PC 9, . . . , PC 34 should be considered noise. The fraction of total variance explained by the seven retained components is obtained from Equation (4) as seen below.

$$\text{Percentage variance explained} = \sum_i \lambda_i \Big/ \sum_n \lambda_n \tag{4}$$

$$= 25.23/34.00$$

$$= 74.2.$$

Table 1. Statistics for the first eight unrotated principal components of the monthly SO_4^{2-} concentration data

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8
Eigenvalue	15.27	2.36	2.01	1.93	1.31	1.23	1.12	0.94
% Var. exp.	44.9	7.0	5.9	5.7	3.9	3.6	3.3	2.8
Cumul. %	44.9	51.9	57.8	63.4	67.3	70.9	74.2	77.0

The elements of each eigenvector were then multiplied by the square root of the associated eigenvalue to obtain the component loadings. These loadings represent the correlations between the component and the station. The square of a component loading indicates the proportion of variance at the individual station that can be attributed to that component. The sum of these squared loadings therefore indicates the total variance accounted for by the component, which as discussed earlier, is called the eigenvalue and is derived from:

$$\lambda_i = \sum_j L_{ij}^2, \tag{5}$$

where L_{ij} is the loading for station j on component i.

Another useful parameter that can be derived from the component loadings is the communality, which is simply the sum of the squared loadings for a particular station, given by:

$$h_j^2 = \sum_i L_{ij}^2. \tag{6}$$

The communality is thus the proportion of the variance for each station accounted for by all of the retained components; the square of each loading being the proportion of the variance accounted for by the particular component. Communalities for each of the 34 stations used in this analysis are provided in Table 2. They range from 0.517 for Bennett Bridge, NY to 0.841 for Oxford, OH, indicating that 51.7% and 84.1%, respectively, of the total variance of the SO_4^{2-} concentrations in precipitation can be accounted for by the seven retained principal components. As expected, the average communality over all 34 stations equals 0.742, confirming Equation (4).

Table 2. Communalities for each ADS site

ADS Site no.	Station	Communality
013A	Lewes, DE	0.767
017A	Georgia St., GA	0.679
020B	Illinois, IL	0.650
022A	S. Illinois U., IL	0.733
025A	Indiana Dunes, IN	0.776
028A	G. Smokey Mts., TN	0.741
030A	Greenville St., ME	0.828
031A	Douglas Lake, MI	0.708
032A	Kellogg, MI	0.778
036A	Meridian, MS	0.755
039A	Hubbard Brook, NH	0.767
041A	Chautauqua, NY	0.791
043A	Whiteface, NY	0.758
044A	Ithaca, NY	0.826
046A	Bennett Bridge, NY	0.517
048A	Brookhaven NY	0.620
050A	Coweeta, NC	0.769
051A	Piedmont St., NC	0.781
052A	Clinton St., NC	0.730
053C	Raleigh, NC	0.777
055A	Delaware, OH	0.746
056A	Caldwell, OH	0.755
057A	Oxford, OH	0.841
058A	Wooster, OH	0.679
064A	Leading Ridge, PA	0.814
072A	Virginia St., VA	0.692
075A	Parsons, WV	0.737
076A	Trout Lake, WI	0.691
077A	Spooner, WI	0.734
150A	Turner Falls, MA	0.736
154A	Rockport, IN	0.782
156A	Ft. Wayne, IN	0.817
164A	Bridgton, ME	0.834
171A	Walker Branch, TN	0.629

SPATIAL ANALYSIS

When principal component loadings are spatially mapped onto their respective stations for each component, isopleths of component loadings can be drawn. Physical interpretation of the non-rotated principal components should be attempted with some degree of caution, because as noted by Buell (1975), the shape of the boundary area to which the correlation coefficients belong is often a factor in determining the topography of the non-rotated components. Because of this, non-rotated principal component analysis is used mainly for data reduction and can be somewhat misleading when used for spatial delineation.

Since one of the main objectives of this paper is to define areas of homogeneous SO_4^{2-} precipitation characteristics, a rotation was performed on the principal components to better segregate the areas that have similar concentration characteristics. Of the two major types of rotation available, oblique and orthogonal, an orthogonal method developed by Kaiser (1958) was selected for this study, because it rigidly rotates the predetermined principal components, while retaining the constraint that the individual components remain orthogonal or uncorrelated to each other. This method increases the segregation between component loadings, which in turn better defines a distinct grouping or clustering of intercorrelated data, thereby making spatial interpretation easier (Horel, 1981).

Comparison between the statistics associated with the seven unrotated components and the seven rotated components, as seen in Table 3, reveals that the total variance explained remains the same, 74.2%, but that the distribution of the variance explained over the individual components is much more uniform.

The rotated principal component analysis allows us to divide the study area into seven separate, contiguous subregions, which each exhibit unique SO_4^{2-} concentration characteristics. As depicted in Fig. 2, loadings from the first rotated component define an area encompassing most of the northeastern states, while the southern Appalachian and mid-Atlantic regions are defined by the second component. The highest loadings from the third component define the southern Great Lakes area, with the southern part of the study area being defined by the fourth component. Loadings from the fifth component define the lower

Table 3. Statistics for the first seven rotated principal components of the monthly SO_4^{2-} concentration data

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Eigenvalue	6.99	4.74	2.99	2.90	2.78	2.60	2.23
% Var. exp.	20.5	13.9	8.8	8.5	8.2	7.6	6.7
Cumul. %	20.5	34.4	43.2	51.7	59.9	67.5	74.2

Ohio River Valley area, while sections of the upper Ohio River Valley are defined by the sixth component. The highest loadings from the seventh and final rotated principal component define the upper Great Lakes region. All seven maps depicted in Fig. 2 can be combined into one map by plotting the highest loadings for each station and the respective rotated principal component. The resulting Fig. 3 depicts the seven separate, contiguous subregions, which each exhibit statistically unique SO_4^{2-} concentration characteristics.

An investigation into the probable underlying physical mechanisms responsible for the statistically unique SO_4^{2-} concentration patterns should intuitively begin with an examination of possible causal relationships with major SO_x emissions areas. In order to establish such a causal relationship, the seven homogeneous subregions as delineated in Fig. 3, have been

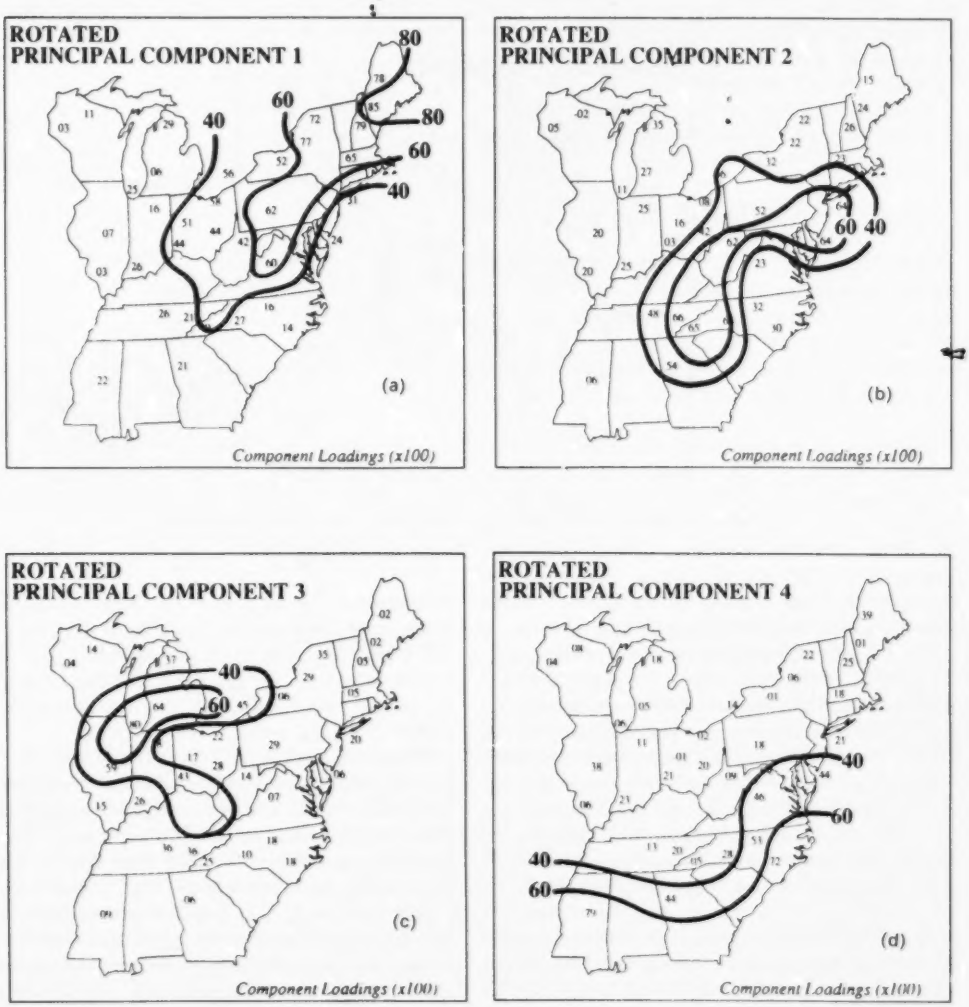


Fig. 2(a–d).

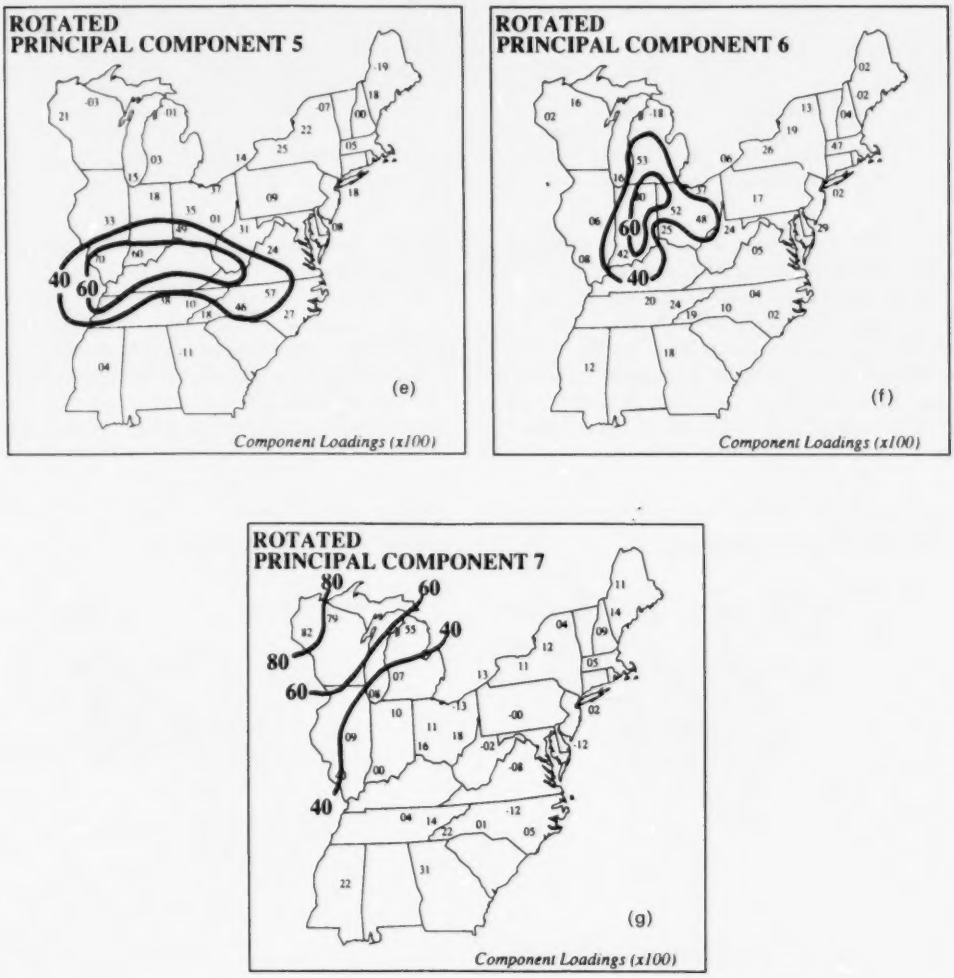


Fig. 2. Component loadings of the seven rotated principal components.

superimposed onto Fig. 4, which depicts gridded annual emissions for 1980 (Doty, 1986).

Examination of this figure reveals several possible relationships. First, with only a few exceptions, the rotated principal components delineate, quite well, the areas of major emissions. For instance, the area defined by the second rotated component corresponds well with an area of moderate SO_x emissions that stretches from northern Alabama and Georgia into western North Carolina. This rotated principal component also includes the small area of emissions found along the Delaware and New Jersey coastline. The third rotated principal component also corresponds fairly well with another area of moderate emissions that stretches from northern Illinois, through northern Indiana into southern Michigan. And finally, the fifth and sixth rotated principal components coincide well with two areas of large SO_x emissions. The fifth encompasses the western Ohio River Valley area where large emissions are found from western Kentucky and Tennessee through southern Illinois and Indiana into western Ohio. The sixth rotated principal component encompasses the large emission areas of eastern Ohio and western West Virginia.

Delineated equally well by the first, fourth and seventh rotated principal components are three relatively 'clean' areas. The first of these encompasses the New England area, while the fourth and seventh encompass southern areas of the study and the upper Great Lakes area, respectively. With only a few exceptions, no major SO_x emissions sources are found within the areas described by these three components.

Such a good correspondence between the emissions areas and the seven subregions as defined by the rotated principal components lends itself to the following causal hypotheses: (1) emissions are likely to be

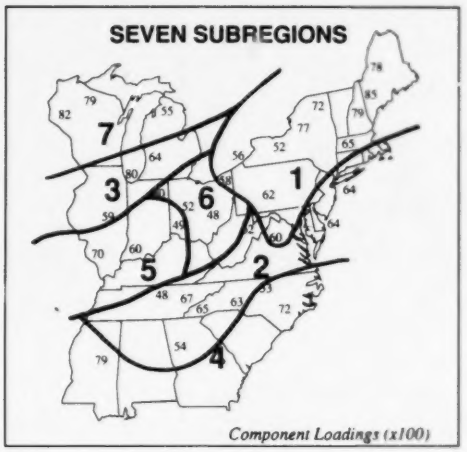


Fig. 3. Seven homogeneous SO_4^{2-} concentration regions as depicted by maximum component loadings.

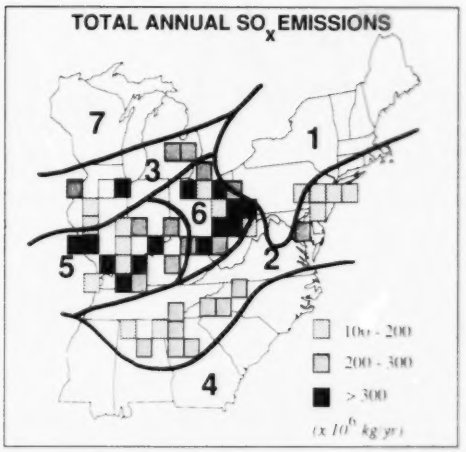


Fig. 4. Seven homogeneous subregions superimposed upon total annual SO_x emissions for 1980.

the major deterministic factor in defining the seven homogeneous subregions; (2) regionally emitted (within subregion) sources are probably responsible for a majority of the high SO_4^{2-} concentration periods that occur in subregions 5 and 6; (3) transport from the major emission sources in other areas is probably responsible for most of the high SO_4^{2-} concentrations periods in subregions 1, 4 and 7; and finally (4) some combination of regionally emitted sources and sources transported from the major emission areas is probably responsible for the high SO_4^{2-} concentrations observed in subregions 2 and 3.

TEMPORAL ANALYSIS

Having delineated seven identifiable SO_4^{2-} concentration patterns, examination of the temporal characteristics of these regions can now be accomplished. Examination of the temporal variance associated with the seven subregions, rather than the 34 individual stations, which are often erratic and noisy due to the small number of contributing observations, results in a smoother and more meaningful temporal pattern.

Initially, the principal component analysis replaced 34 stations, measured over 72 months, with seven principal components having no temporal measurement. By introducing the principal component score (PC_s), a derivation of similar temporal measurements for the principal components over the same 72 months can be achieved.

The principal components are identified in terms of the original stations, the larger the loading the more important the station is in the interpretation of the component. Therefore, if a month has high values for the stations with large loadings, then it should have a large value on that component. The PC score for month m on component i are designed to meet this requirement and are defined by Johnston (1978) as follows:

$$(PC_s)_{mi} = \sum_j O_{mj} L_{ji}, \tag{7}$$

where O_{mj} is the observation for month m on station j, L_{ji} is the loading of station j on component i, and the PC_s is the component score for month m on component i and is summed over all 34 stations. The component scores have been standardized, therefore they have a mean of zero and a standard deviation of 1.

When plotted as a time series over the entire period of record, as depicted in Fig. 5, the PC_s provide excellent insight into the spectrum of temporal variance experienced by each of the subregions. In order to better illustrate any general or seasonal trends that may be present in the data the PC_s have also been smoothed using a cubic spline function. The ordinate of each plot represents standard deviations from the mean, while the abscissa represents the 72 month period of record.

Examination of the smoothed PC_s associated with the seven subregions reveals an in-phase anomaly pattern, as it is apparent that most of the time series generally follow a similar annual cycle, in which we see a propensity for the highest SO_4^{2-} concentration periods to occur during the summer, while the lowest values tend to occur during the winter. It should be noted that the terms 'high' and 'low' SO_4^{2-} concentration periods are not meant in the absolute sense, but instead refer to the largest positive and negative departures from normal of the standardized PC_s.

This annual cycle is generally most prevalent and smoothest in subregions 1, 2, 3, 4 and 7 (which were earlier defined as either 'clean' subregions or sub-

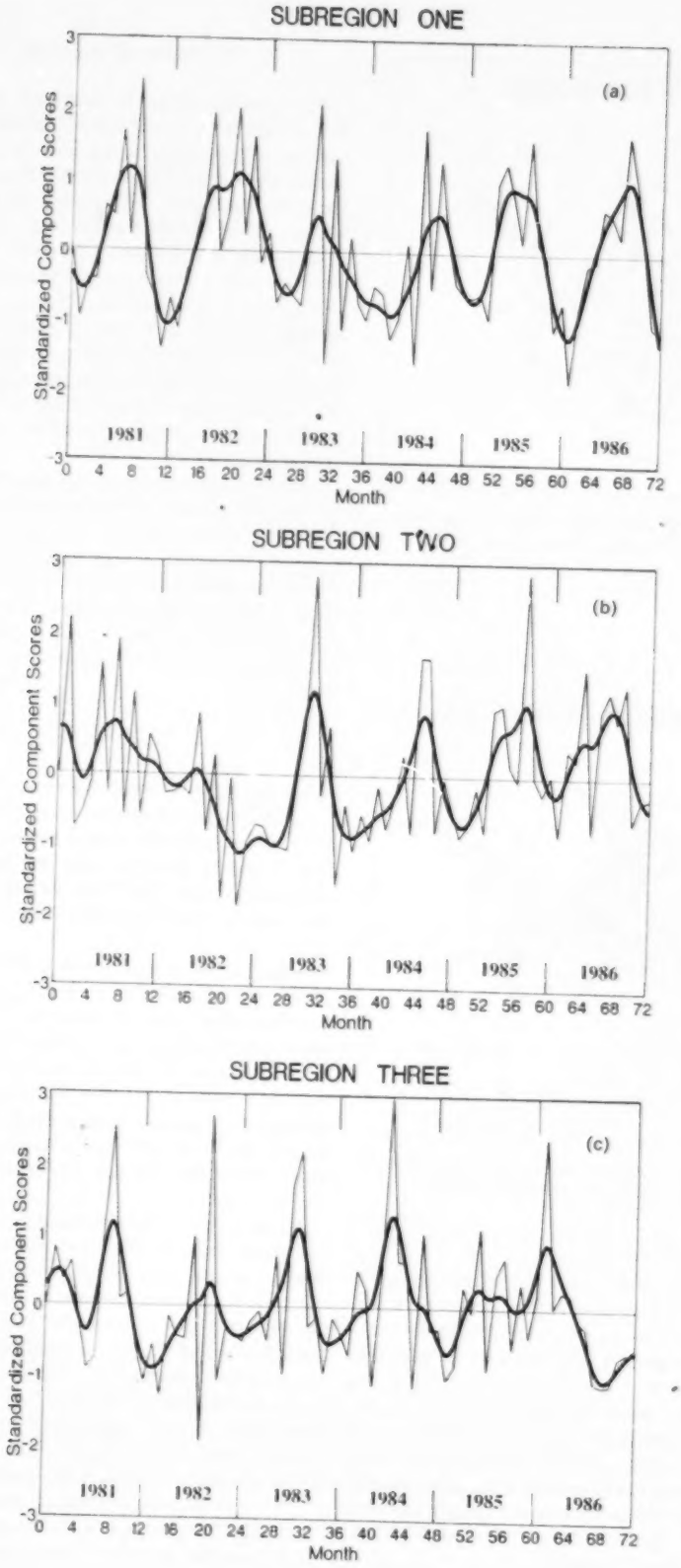

SUBREGION ONE
(a)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month
SUBREGION TWO
(b)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month
SUBREGION THREE
(c)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month

Fig. 5(a–c). —

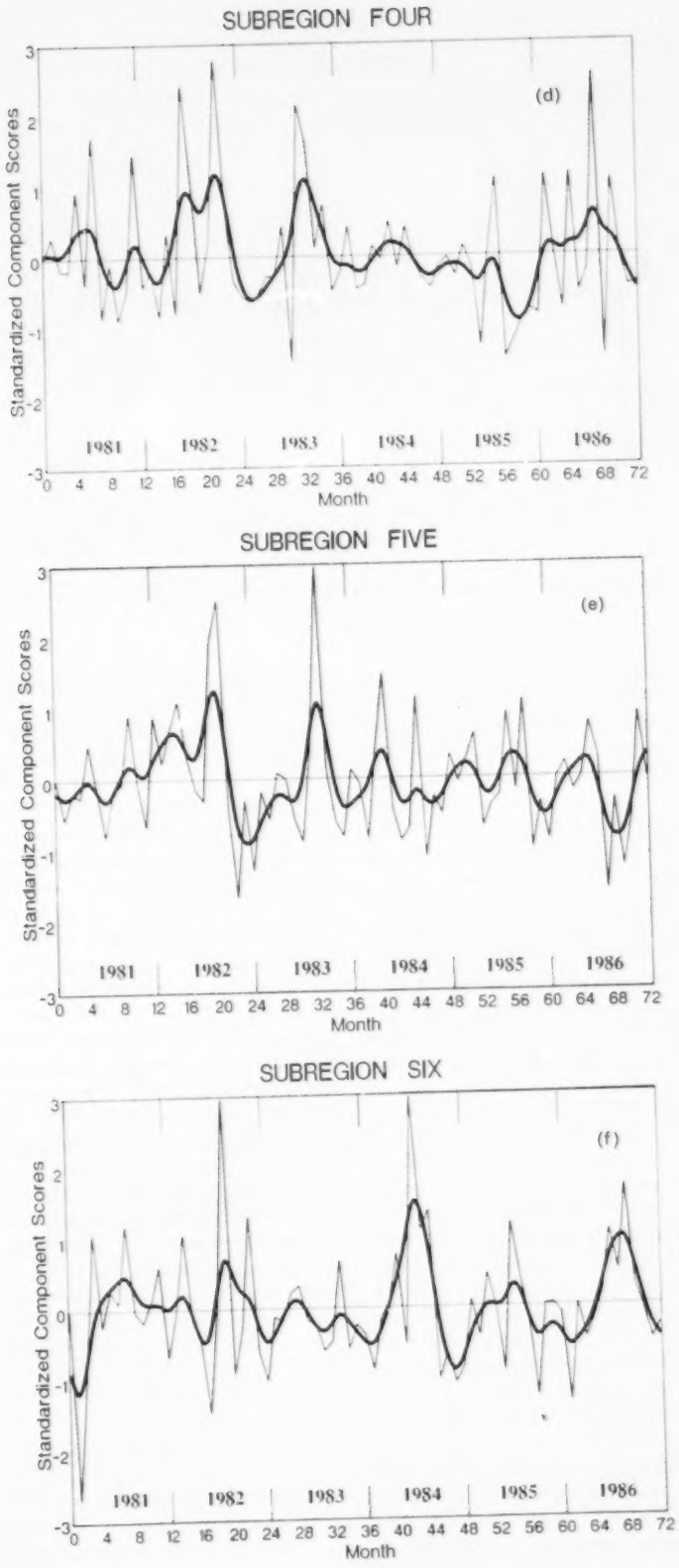

SUBREGION FOUR
(d)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month
SUBREGION FIVE
(e)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month
SUBREGION SIX
(f)
Standardized Component Scores
3
2
1
0
-1
-2
-3
1981
1982
1983
1984
1985
1986
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72
Month

Fig. 5(d–f).

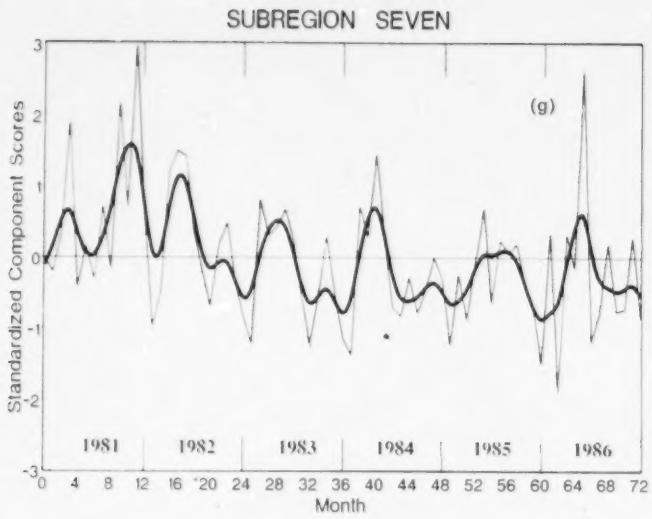


Fig. 5. Time series of the standardized principal component scores associated with the seven subregions.

regions which contained only moderate emissions), where transport is hypothesized to be a major contributor of the SO_4^{2-} concentrations found in the precipitation. The lack of large areas of major, regionally emitted SO_x within these subregions results in a smoother time series, which reflects the smooth seasonal variation of meteorological and chemical processes such as solar radiation, temperature and precipitation type, that control SO_4^{2-} processes.

This annual cycle is, however, less prevalent in subregions 5 and 6, which contain large SO_x emissions that are thought to be the major contributor to SO_4^{2-} concentrations within these subregions. The influence of these major, regional (within subregion) emissions appear to dominate these time series as the annual cycle is almost completely obscured by frequent perturbations.

Examination of the raw PC_s reveals another interesting and somewhat unusual phenomenon in that there is a tendency for low concentration periods to occur in groups of 4 or 5 months, while most of the high concentration periods are only sustained for 1 month. This characteristic is illustrated by the more spiked nature of the positive anomalies, as compared to the somewhat smoother nature of the negative anomalies. This phenomenon may indicate that the natural, meteorological processes responsible for high SO_4^{2-} concentration periods cannot generally be sustained for periods longer than 1 month.

These periods of high concentration are best illustrated by examining all months for which the PC_s exceeded 2.0 standard deviations (hereby abbreviated by s.d.) above the mean. Examination of the time series associated with subregion 1 reveals that the PC_s exceeded 2.0 s.d. 3 times during the 72 month period: once during August 1981, and again during August 1982 and during June 1983. The PC_s associated with subregion 2 also exceeded 2.0 s.d. 3 times. The first was during January 1981 (which is one of rare the winter periods), while the remaining two occurred during July 1983 and September 1985. Subregion 3 experienced more (five) extreme positive values than all of the other subregions. These occurred during August 1981, August 1982, July 1983, June 1984 and January 1986. PC_s associated with subregion 4 exceeded 2.0 s.d. 4 times; during May and September of 1982, and during July of 1983 and 1986. The next two subregions only exceeded 2.0 s.d. twice, during August of 1982 and 1983 for subregion 5 and during July 1982 and June 1984 for subregion 6. And finally, the PC_s associated with subregion 7 exceeded 2.0 s.d. above the mean 3 times; during September and November of 1981 and again during May 1986.

Examination of the meteorological conditions associated with each of these high concentration periods (as well as for those associated with the low concentration periods) may lead to a better understanding as to the physical reasons for such events. Any statistical correlations or physical linkages found between the time series data and other meteorology variables may offer an explanation of the synoptic scale physical mechanisms underlying these relationships.

CONCLUSIONS

The spatial and temporal variability of SO_4^{2-} precipitation concentrations over the eastern U.S. during the period 1981–1986 was examined through the use

of a multivariate statistical technique called Principal Component Analysis (PCA). Due to the copious amount of data resulting from such a large scale investigation and because such data tend to be erratic or noisy, it was necessary to employ an analysis method which would identify, through a reduction in data, the recurring and independent modes of variability within the larger, noisy data set. Such analysis was achieved through the use of PCA, which allowed for the reduction of the original data set, which contained 34 correlated variables (the stations), down to seven uncorrelated variables (the principal components), while still explaining 74.2% of the total variance.

Application of Kaiser's Varimax orthogonal rotation led to the identification of seven separate, contiguous subregions which each exhibit unique SO_4^{2-} concentration characteristics. These subregions were found to corresponded well with major SO_x emission patterns, leading to a hypothesis of causality in which proximity to emission areas was the deterministic factor in defining the seven homogeneous subregions. Regionally emitted sources are thought to be responsible for a majority of the high SO_4^{2-} concentration periods that occurred in subregions 5 and 6; whereas transport from major emission sources is thought to be responsible for high concentrations in subregions 1, 4 and 7, and finally a combination of regionally emitted and transport emissions is thought to be responsible for the high concentrations found in subregions 2 and 3.

Examination of the principal component scores associated with each of the seven subregions supported the hypothesis of causality and revealed a general annual cycle in which the occurrence of high concentrations is much more likely during the summer months than the winter months. This seasonality was most prevalent in subregions 1, 2, 3, 4 and 7, where it can be related to be seasonal variations of the meteorological and chemical processes that control SO_4^{2-} processes; and it was least prevalent in subregions 5 and 6, where influence from regionally emitted sources obscure the seasonality.

In order to further substantiate the hypothesis that causality exists between the emissions and the seven contiguous subregions, it is recommended that a variety of interpretative studies, including trajectory analysis, be undertaken by the extended research community. Upon detailed examination and physical interpretation, such analyses should lend themselves toward furthering our understanding of the source/receptor relationships of acid precipitation.

Disclaimer—Although the research described in this article has been supported by the United States Environmental Protection Agency, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

REFERENCES

Ashbaugh L. L., Myrup L. O. and Flocchini R. G. (1984) A principal component analysis of sulfur concentrations in the western United States. *Atmospheric Environment* **18,** 783–791.

Buell C. E. (1975) The topography of the empirical orthogonal functions. *Preprints of the Fourth Conference on Probability and Statistics in Atmospheric Sciences*, 188–193.

Crawley J. and Sievering H. (1986) Factor analysis of the MAP3S/RAINE chemistry network: 1976–1980. *Atmospheric Environment* **20,** 1001–1013.

Dana M. T. and Easter R. C. (1987) Statistical summary of event precipitation chemistry from the MAP3S network, 1976–1983. *Atmospheric Environment* **21,** 113–128.

Doty K. G. (1986) Spatial patterns of NADP sites in the eastern United States for specific ions using cluster analysis. *Preprints of the Second International Specialty Conference on Meteorology of Acidic Deposition*, 365–380.

Gatz D. F. (1978) Identification of aerosol sources in the St. Louis area using factor analysis. *J. appl. Met.* **17,** 600–608.

Hayden B. P. and Smith W. (1981) Season-to-season cyclone frequency prediction. *Mon. Wea. Rev.* **110,** 239–253.

Hendry C. D. and Brezonik P. L. (1982) Chemistry of precipitation at Gainesville, Florida. *Envir. Sci. Technol.* **14,** 843–849.

Henry R. C. and Hidy G. M. (1979) Multivariate analyses of particulate sulfate and other air quality variables By principal components—Part I. Annual data from Los Angeles and New York. *Atmospheric Environment* **13,** 1581–1596.

Hopke P. H., Gladney E. S., Gordon G. E., Zoller W. H. and Jones A. G. (1976) The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. *Atmospheric Environment* **10,** 1015–1025.

Horel J. D. (1981) A rotated principal component anlysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Mon. Wea. Rev.* **109,** 2080–2092.

Johnston R. J. (1978) *Multivariate Statistical Analysis in Geography*. Longman Press, London.

Kaiser H. F. (1958) The Varimax criterion for analytical rotation in factor analysis. *Psychometrika* **23,** 187–201.

MAP3S/RAINE Research Community (1982) The MAP3S/RAINE precipitation chemistry network: statistical overview for the period 1976–1980. *Atmospheric Environment* **16,** 1603–1631.

Overland J. E. and Preisendorfer R. W. (1982) A significance test for principal components applied to a cyclone climatology. *Mon Wea Rev.* **110,** 1–4.

de Pena R. G., Walker K. C., Lebowitz L. and Micka J. G. (1985) Wet deposition monitoring—effect of sampling period. *Atmospheric Environment* **19,** 151–156.

Richman M. B. (1981) Obliquely rotated principal components: an improved map typing technique? *J. appl. Met.* **20,** 1145–1159.

Seilkop S. K. and Finkelstein P. L. (1987) Acid precipitation patterns and trends in eastern North America, 1980–1984. *J. clim. appl. Met.* **26,** 980–994.

Sisterson D. L., Wurfel B. E. and Lesht B. M. (1985) Chemical differences between event and weekly precipitation samples in Northeastern Illinois. *Atmospheric Environment* **19,** 1453–1469.

Smeyers-Verbeke J., Den Hartog J. C., Dekker W. H., Coomans D., Buydens L. and Massart D. L. (1984) The use of principal component analysis for the investigation of an organic air pollutants data set. *Atmospheric Environment* **18,** 2471–2478.

Thomas W. (1986) Principal component analysis of trace substance concentrations in rainwater samples. *Atmospheric Environment* **20,** 995–1000.

Watson C. R. and Olsen A. R. (1984a) North American Acid

Deposition System (ADS) and its use in spatial analysis. *Preprints of the* 77th *Air Pollution Control Association Annual Meeting*.

Watson C. R. and Olsen A. R. (1984b) Acid Deposition System (ADS) for reporting statistical reporting—system design and user's code manual. U.S. Environmental Protection Agency Report No. EPA-600/8-84-023, Research Triangle Park, NC.

Wolff G. T., Siak J., Chan T. L. and Korsog P. E. (1986) Multivariate statistical analyses of air quality data and bacterial mutagenicity data from ambient aerosols. *Atmospheric Environment* **20**, 2231–2241.